



# Artificial Intelligence Living Lab (SEDIA) Final Event

November 27, 2023

At Spanish Foundation for Science and Technology (FECYT)





# Technological Breakthroughs

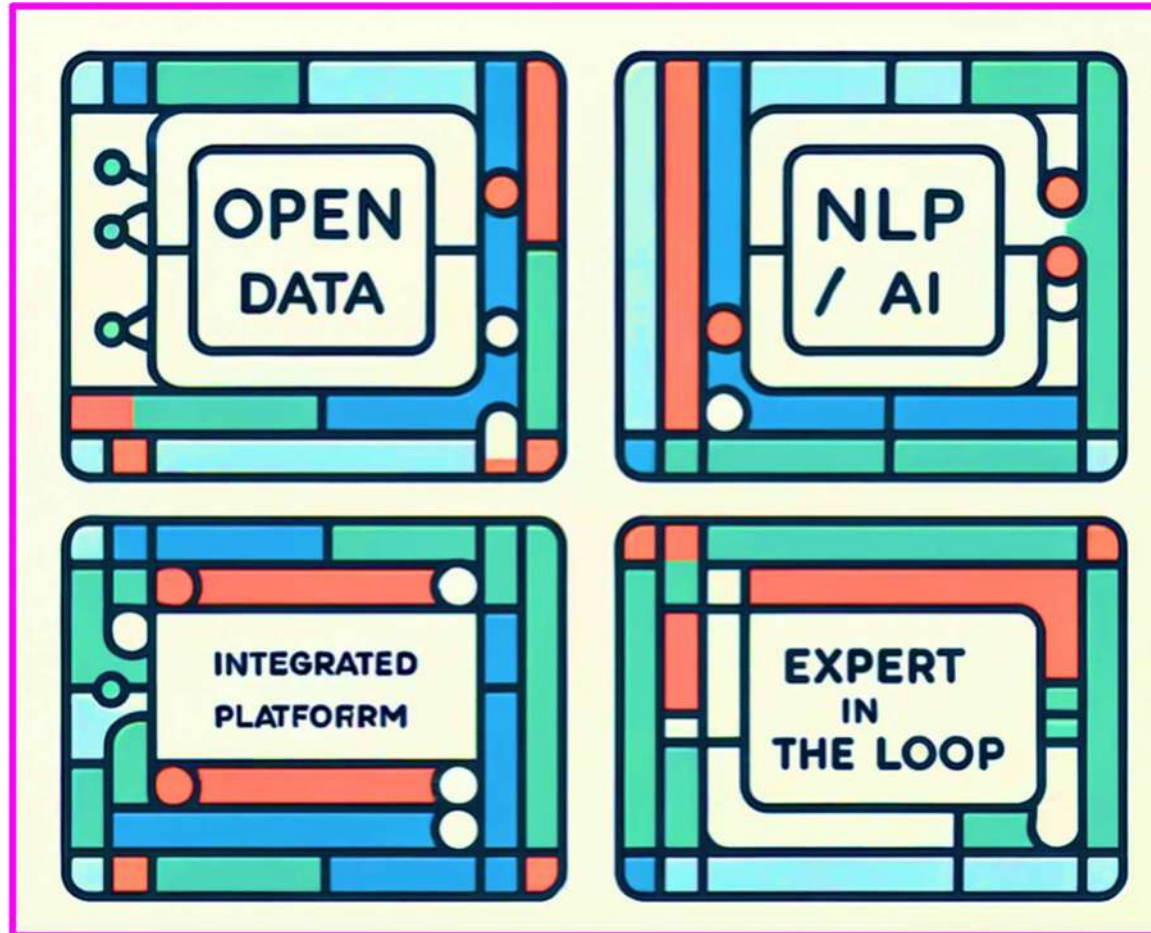
Jerónimo Arenas-García (Universidad Carlos III de Madrid – UC3M)

November 27, 2023

AI Living Lab (SEDIA) Final Event



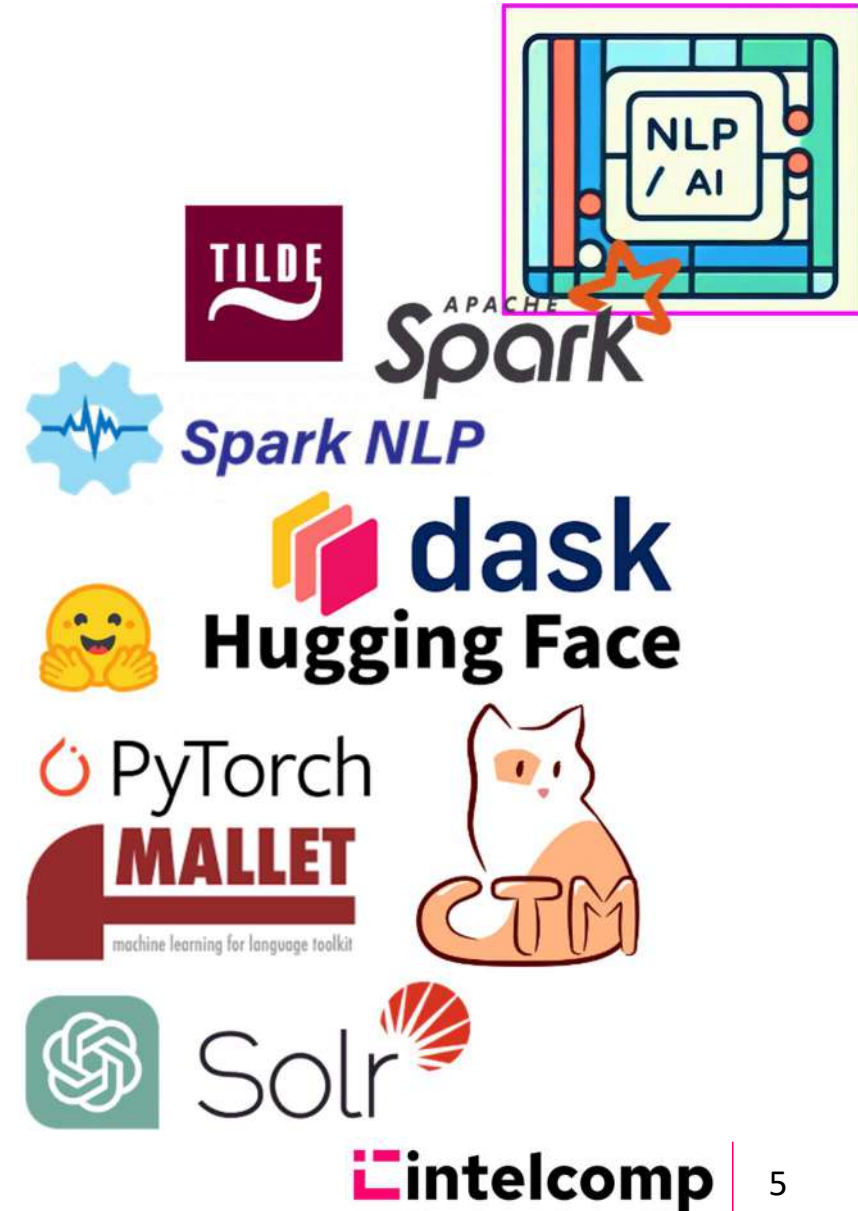
## Technological challenges: Cornerstones of the project



- IntelComp's data lake collects several STI-related open data sets
  - Metadata heterogeneity
  - Variable curation quality
  - Very large datasets
- IntelComp's approach:
  - Unstructured datalake
  - Distributed storage (HDFS, parquet)
  - English as an anchor language
  - Text-based representations

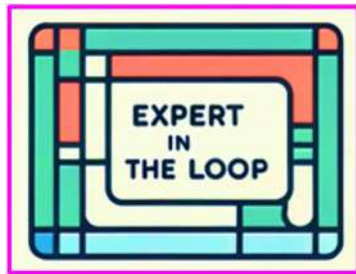
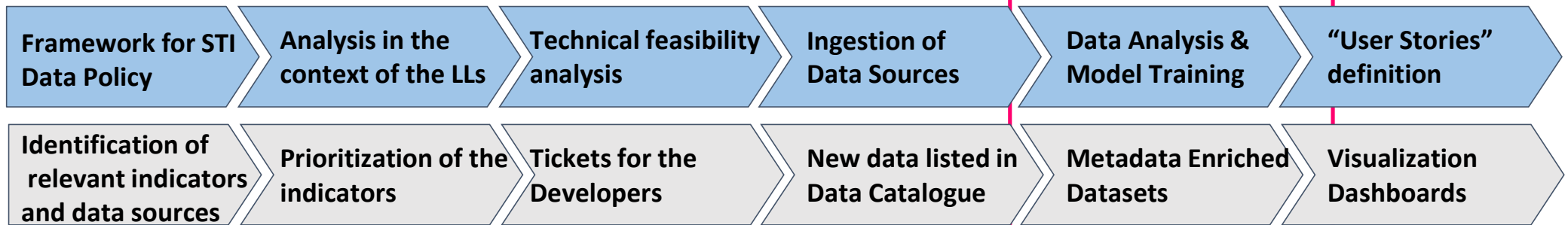


- Scalable NLP and NMT
  - GPU processing
  - Spark NLP for parallel processing
- Text-based AI
  - Supervised classification
  - Zero-shot classification
  - Document domain selection
  - Topic Modeling
- LLMs
  - Transformer-based document representations
  - GPT models to improve interpretability





## THEORETICAL WORK



### Platform Functionalities:

- Data Ingestion and Preprocessing
- Data Enrichment (AI model training and inference)
- User services (AI model exploitation)

# IntelComp Tools

**Find a dataset**

Which dataset are you looking for?

SEARCH

**Data Catalogue**

**Data Services / APIs**

The RESTful API allows the user search engine for efficient data storage and retrieval of digital corpora and their associated metadata. Data is formatted according to the specifications provided by the application, enabling seamless integration and shared capabilities. The API also offers a range of query options to explore information retrieval.

Collections: Search, Sub-corpora, Models, Queries

**Corpora** (Cross-linked corpora (i.e., interlinked corpora))

- /corpora/tdbsearchhead/tdb/
- /corpora/deletecorpus/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/
- /corpora/tdbsearchhead/tdb/

**Models** (Models related corpora (i.e., interlinked models))

**Queries** (Specific for queries)



**Start**

Here is an overview of IntelComp's Model Maker

- Corpora**  
Use this space to select a corpus from the Data Lake to be used with the training of a model. You can also create a new corpus by merging other available ones or select a subcorpus based on a category.
- Models**  
Use this space to select a model to work with create a new model, curate a model or manage the available models.
- Stopwords & Equivalences**  
Use this space to manage your lists of stopwords and equivalences or create new ones.

**IMT**

**STI Viewer**

66134 Publications  
2279781 Citations  
2166 Publications with Citations

Scientific Production Trends  
Open Access publications in Cancer over time

**EWB**

Cancer: 0-Overexpressed Nucleic Acids, 0-Underexpressed Nucleic Acids, 0-Overexpressed Proteins, 0-Underexpressed Proteins

High Negative Breast Cancer is the most aggressive among breast cancer subtypes. The conventional treatment strategy is ineffective in about 50% of cases, with patients presenting a higher risk of recurrence and metastasis. The failure of the currently available chemotherapeutic strategy is also due to the lack of cancer-practical models that are "close to reality" and therefore able to accurately reproduce the tumor microenvironment, the transcriptional reprogramming and the cancer cells plasticity of patients with poor outcomes. Currently, mRNAs and miRNAs are believed to be the components of all major signal pathways and involved in the transcriptional reprogramming that occurs after TNBC treatment with more advanced therapies. However, transcriptional reprogramming in vitro models cannot study the molecular mechanisms underlying adaptive responses to chemotherapy by identifying transcriptional phenotypes that somehow make tumor cells tolerant to drugs, without requiring a substantial growth reduction. The CONTACT project aims at developing a 3D transcriptional TNBC in vitro model able to shed light on the mechanisms of tumor cells reprogramming that trigger drug tolerance. The model will include several components along with cancer cells and represents an important tool involving the role of the stroma in transcriptional reprogramming. Transcriptional 3D breast cancer in vitro models will accelerate the establishment of personalized treatment protocols that take into account what exactly, in each patient, leads to transcriptional reprogramming. The CONTACT project will boost the scientific career of the experimental researcher and will lead to an independent risk in research. The project will also benefit the expertise of the experimental researcher in the area of cancer genetics. Her expertise in 3D tumor models will open the way for a new research line in the scientific potential of the best institutions.

**Graph Explorer**

**intelcomp**

Welcome to IntelComp's policy participation portal

**PP**

## User-in-the-loop model training **IMT**

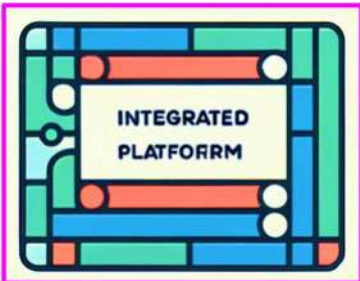
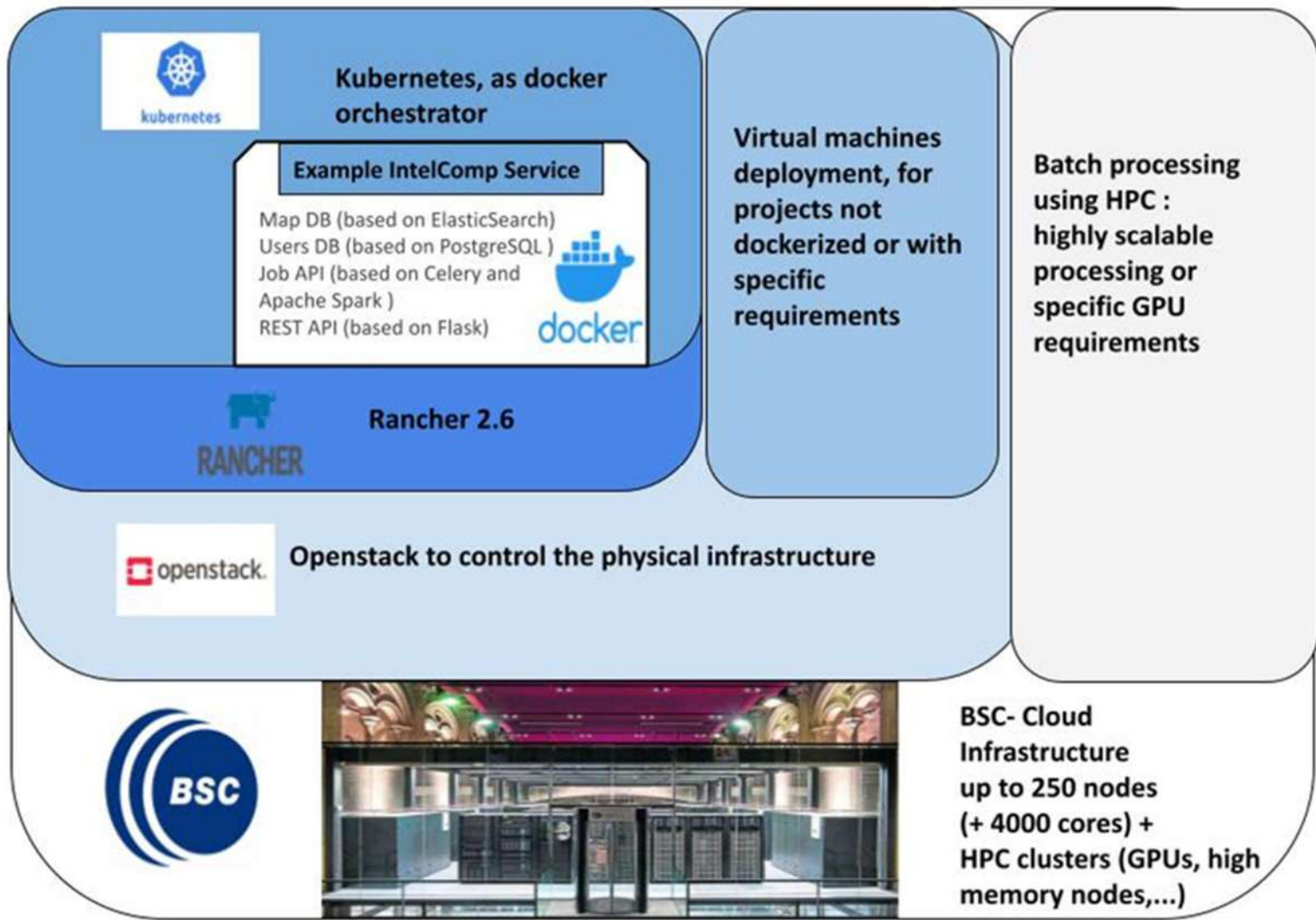


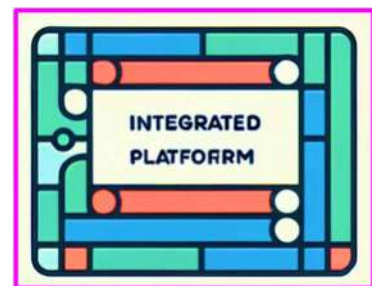
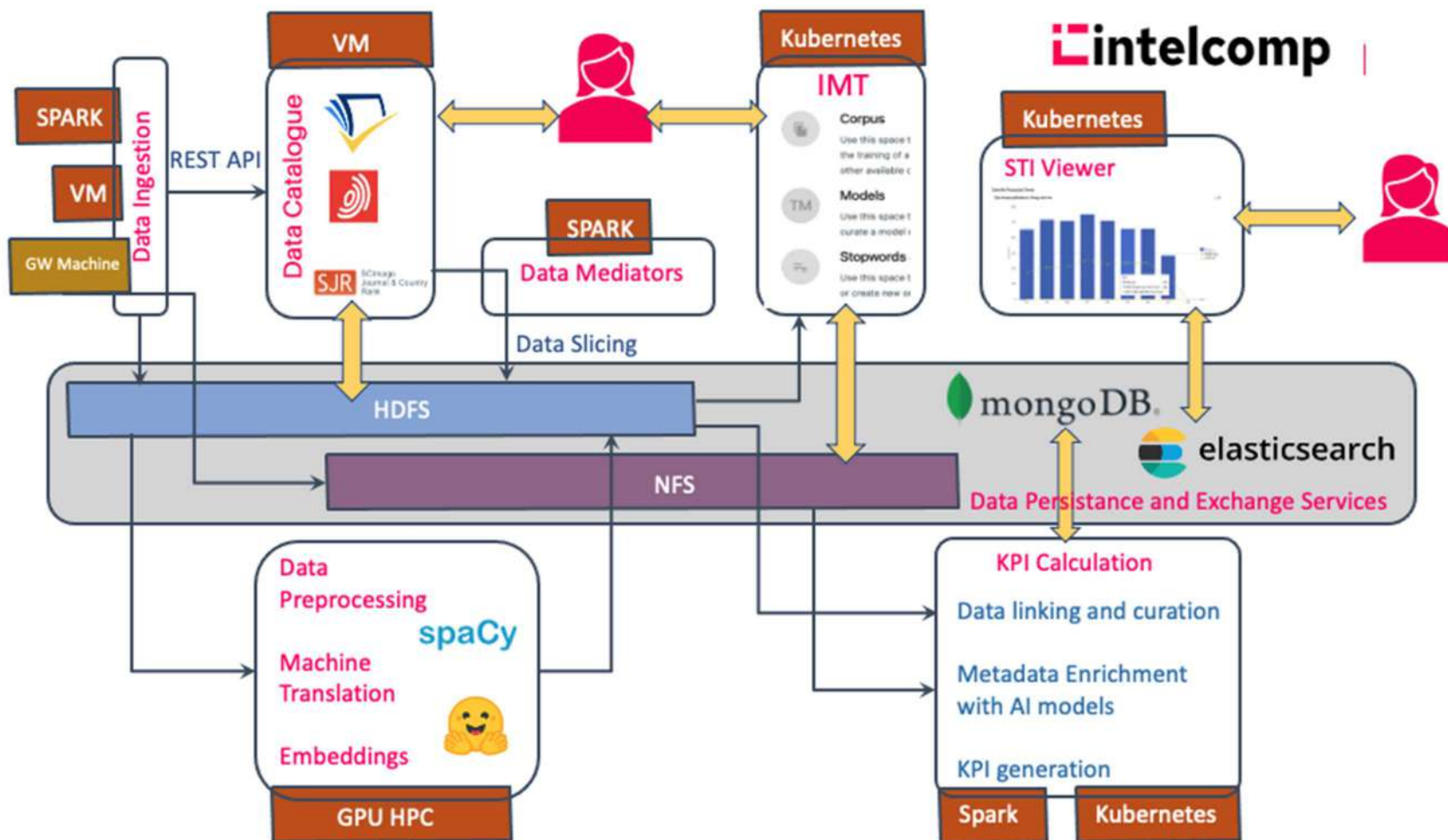
- User-curated domain classification
  - Restrict the analysis to the scope of interest
  - Better similarity metrics for the selected domain (=better services in the EWB)
- User-curated topic models
  - Implement a dimension for cross-comparison among heterogeneous datasets
  - Visualize this information in the STI Viewer
- Create your own classification models
  - Enrich datasets with the taxonomies that are relevant for you

The goal is better experience with the STI Viewer and the EWB

- Maximize alignment with domain experts' intuition
- Improve recommendations of the EWB







intelcomp



# Main achievements of the IntelComp Project

Joseba Sanmartin (Spanish Foundation for Science and Technology – FECYT)

November 27, 2023

AI Living Lab (SEDIA) Final Event



## I. A reminder: The IntelComp platform

- A platform to improve and automatise the tasks of managers of R&D programs at ministries (e.g. SEDIA) and funding agencies, by exploiting millions of documents with text analytics.
- Digital records of all STI activities & Rapid evolution of automatic text analysis: ideal conditions for application of text tools in science and innovation policy.
- **Public Administrations need a platform like IntelComp to exploit the opportunity of text as data** to get insights that previously were not possible and to automatise some tasks.

## II. Main achievements: A theoretical framework for text analysis in STI policy

- Our framework combines the functions of the innovation system with the stages of the policy cycle. By crossing these dimensions, we got questions of interest to policymakers.
- We focused on the stages where text analytics can be more useful:
  - Agenda setting or intelligence gathering
  - Policy formulation
  - Policy implementation
  - Policy monitoring

## II. Main achievements: A theoretical framework for text analysis in STI policy (cont.)

- *Agenda setting.* Program managers can use text as data to analyze the context of the program: focus on specific domains and measure research and innovation activities.
- *Policy formulation.* Program managers can use text analytics to study previous funding in the context of the current research and innovation landscape.
- *Policy implementation.* In the evaluation of the proposals for funding, call managers can exploit the texts of the proposals to analyze and compare them, and to find suitable reviewers.
- *Policy monitoring.* Monitoring officers can explore the text data associated to the inputs and outputs of programs to follow their execution.

### III. Main achievements: How we gathered the requirements for the platform

- With this theoretical framework in mind, Living Labs have practiced with reliable data to define the users' needs for the different interactive tools.
- Public Administrations can apply IntelComp tools in several stages of the policy process:

Stage of the policy cycle	STI Viewer – STI Policy Participation Portal	Graph Visualizer	Evaluator Workbench
Agenda setting			
Policy formulation			
Policy implementation			
Monitoring of results			

### III. Main achievements: How we gathered the requirements for the platform (cont.)

- *STI Viewer and STI Policy Participation Portal:*
  - The Living Labs have prioritized the list of measurements for the analysis of the context and the description of the results of public funding.
  - The Living Labs have also provided requirements for the usability of the dashboards.
- *Graph Visualizer:*
  - Program managers can learn about the thematic distribution and evolution of funding with this tool.



### III. Main achievements: How we gathered the requirements for the platform (cont.)

- *Evaluation Workbench.*
  - Call managers can analyze set of proposals and documents, search similar documents, classify the proposals, find suitable reviewers and detect conflicts of interest.
  - The Greek funding agency has defined the requirements and provided the data to develop this tool for the assessment of proposals.
- *Interactive Model Trainer.*
  - Domain experts have defined the requirements of this tool, which guarantees a human-in-the-loop approach in the text analysis.
  - These models feed the other tools of the platform. For instance, in the AI Living Lab experts in AI have co-created the models used to calculate the indicators shown in the STI Viewer.

#### **IV. Main achievements: A working platform**

- All the services and tools are integrated and deployed.
- The IntelComp tools are sufficiently mature: ministries and funding agencies can apply them in use cases of any domain.



# IntelComp Business Models

## Main aspects

Daniel Jimenez (NTT DATA)

November 27, 2023

AI Living Lab (SEDIA) Final Event



# IntelComp solution: modular and open-source components

## IntelComp Integrated Applications, a modular solution

### IntelComp Catalogue



The IntelComp Catalogue offers a single store of data collections to be exploited by the IntelComp tools. It allows the processing of a variety of documents, unifying them and enriching the information to obtain improved and more complete results.

[Visit service](#) →

### STI Viewer



An interactive AI-assisted tool to understand the current STI landscape and visualize its evolution in several domains and countries, observe emerging topics, detect the main players, and learn about the distribution and results of public funding. The target audience of this tool are STI policy analysts who can explore and visualize a comprehensive set of indicators for the agenda setting and the monitoring & evaluation stages of the STI policy cycle.

[Visit service](#) →

### Policy Participation Portal



The Participation Portal helps to facilitate the integration of diverse stakeholder input into STI policy-making processes. It offers a user-friendly interface where opinions are collected through surveys enriched with data visualizations from the STI Viewer. This tool fosters a collaborative and transparent policy-making environment ensuring that decisions are not only data-driven but also responsive to societal needs and industry challenges. Academia, Industry, and Citizen representatives are the target audience of this tool.

[Visit service](#) →

### Evaluation Workbench



A tool that assists call managers at STI funding agencies in the management of proposals for funding. They can analyze set of proposals and documents, search similar documents, automatically classify the proposals, find suitable reviewers and detect conflicts of interest. Evaluation Workbench makes the evaluation process easier saving time and increasing efficiency. This tool is intended to public administration officials who manage proposals.

[Visit service](#) →

### Interactive Model Trainer



This tool is intended for experts in a field. Through a set of interactive services available in the Interactive Model Trainer (IMT), it is possible to incorporate the knowledge of experts in specific domains during model training, thus obtaining higher quality and more interpretable topics. IMT also allows experts to play an active role in the creation and validation of these models. These models are then used by the end-user tools (the STI Viewer, the STI Participation Portal, the Graph Explorer and the Evaluation Workbench), who can customize their analyses, comparisons, and visualizations according to the new models trained in the IMT.

[Visit service](#) →

### Graph Explorer



Graph Explorer is an interactive AI-assisted tool to understand the current STI landscape and its evolution, explore thematic communities, and learn about the thematic distribution and evolution of STI funding. The target audience of this tool are STI policy analysts who can explore large graphs of millions of documents for the agenda setting and the monitoring & evaluation stages of the STI policy cycle. It is a very useful tool when you have a large amount of data to analyze and you need insights at aggregated but also granular levels.

[Visit service](#) →

All IntelComp components are **open** and **available** in GitHub



# IntelComp target customers and value offering

The **typical user** of Intelcomp tools is a **STI policy analyst** from two types of organisations:

- 1) **Funding Agencies**
- 1) **Policy Analysis Firms** (providing services to Funding Agencies)

## IntelComp value offering

- 1) **Relevant savings in extracting insights**, moving from a highly manual process to a **semi-automated end-to-end process designed for STI policy analysis and supported by artificial intelligence**.
- 1) **Global coverage of the different stages of the policy cycle** through the **IntelComp integrated solution** versus specific adhoc activities, **obtaining synergies and efficiencies**.

# IntelComp exploitation scenarios

IntelComp offers **two main approaches**:

## Open scenario

Users can choose the **modules** from **GitHub** and customise/integrate them in their own processes

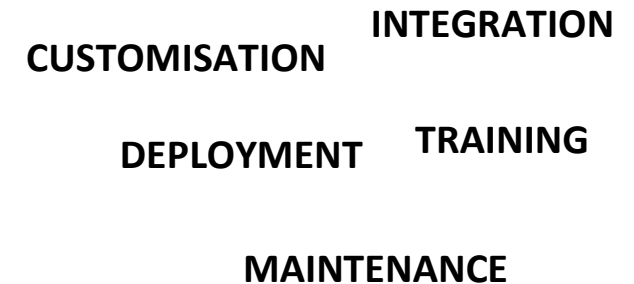


## Scenarios based on IntelComp partners services

Use of the **integrated solution** through on a pay-per-use model or project-based approach



Professional services focused on tailoring IntelComp modules or the integrated solution from partners



The overall solution comprises many different technologies and capabilities, so a customer may not have all the required capabilities to customise the modules on their own. Customers can rely on IntelComp partners for the creation/tailoring of the modules.